Response to comments from Reviewer 4 on "*The evolution of process-based hydrologic models: Historical challenges and the collective quest for physical realism*" by Martyn P. Clark et al.

[Responses are in red font at the bottom each sub-section].

1 General

This manuscript is an interesting contribution to the on-going community debate on how to advance hydrologic models. Given the nature of this manuscript, I will below remark on the three main areas of the commentary, rather than commenting on individual parts. Hopefully my comments will help to expand the nice discussion in this manuscript even further.

One overall issue that might be stressed more in this commentary is that the three areas outlined (model structure, model parameters and model execution) are interdependent, and that improving one requires advancements in the others. For example, it is difficult to reduce parameter spaces for complex models if computing demands do not allow us to explore such spaces thoroughly in the first place.

Yes, good point. We now highlight these interdependencies in the Introduction:

> *We discuss these modeling challenges separately, while recognizing that these modeling challenges are strongly interdependent (e.g., a complex model structure may have large computing demands, restricting the extent to which it is possible to explore alternative model parameter sets). We will touch on these interdependencies in the individual sections of the paper.*

We also expand discussion on the interdependencies among modeling challenges in the individual sections of the paper.

2 Model Structure

One issue to mention here might be that there is a trade-off between our ambition to have models that are flexible enough to produce a high performance when matched against observations, that are parsimonious so that parameter uncertainty is low, and that show a high degree of realism in the sense that they are consistent with reality – often equated with models of higher resolution (Wagener, 2003). These are often seen as conflicting objectives. It might be worth discussing that model realism can be achieved with simpler models, while more complex models can still miss key processes or get key fluxes wrong. One example is the recent paper by Hartmann et al. (2017), which compared a widely used global model (PCR-GLOBE) with a much simpler model (Var-Karst). The latter included subsurface heterogeneity and produced much more realistic recharge estimates for karst regions. Another example might be the lack of preferential flow representation in many otherwise complex models.

So how can we ensure that our models are not missing key processes, while we focus on improving details elsewhere? Maybe the top-down approach discussed in another current commentary by Clark and Hrachowitz is a strategy to approach this problem?

Good point. We now include discussion of model tradeoffs in the section on computing:

> *In exploring these solutions we recognize that there is not necessarily a tradeoff between physical realism and computational efficiency – the linkage between spatial complexity and process complexity may be rather weak, as models run using a large number of spatial elements may miss dominant processes [e.g., Hartmann et al. 2017]*

2.1 Model Parameters

The authors mention the use of signatures for constraining parameter spaces. I think this part might be worth expanding a bit. Such strategies are still not used regularly for distributed models though some nice examples of their value exist. One such example is the paper by Troy et al. (2008) where Eric and his students/colleagues use runoff ratio to constrain VIC at the grid scale. The resulting parameter estimates are much improved by this process.

So, what information can we use to constrain our hyper-resolution models? This information can come from a range of places. For example, it might be possible to synthesise previous experimental and modelling studies that have focused on individual places to gain a better expected value of flux magnitudes across larger domains, or we might be able to use observed vertical fluxes of moisture and energy as 'weak' constraints to account for scale differences between measurement and model scales (e.g. both done by Hartmann et al., 2017, in relation to karst recharge). Or we might be able to regionalize signatures as constraints beyond those relevant for streamflow, but maybe relevant for ET or other fluxes/states (e.g. the Troy et al., 2008, strategy). How can we reduce the acceptable output space of a model to reduce parameter uncertainty?

We have added discussion to the section on model parameters:

> *[...] there is considerable scope to improve the way that multivariate data is used to constrain model parameter values. A key path forward is to identify different signatures from the data that can be used to improve parameter values in different parts of the model [Gupta et al. 2008; Yilmaz et al. 2008; Pokhrel et al. 2012; Vrugt and Sadegh 2013; Rakovec et al. 2015]. For example, Troy et al. [2008] use regionalized estimates of the runoff ratio to constrain the VIC model at the grid scale, and there is much more that can be done using such methods [e.g., see the approach of Yadav et al. 2007]. In the distributed model context, signatures related to energy and moisture fluxes may now be constrained by remote sensing imagery, e.g. of skin temperature or ET, though this strategy is far from common today. Similarly, remotely sensed estimates of surface water levels [Revilla-Romero et al., 2016] and total basin storage [Tangdamrongsub et al., 2015] could be used as well as reported statistics on water withdrawal [Wada et al., 2014].*

2.2 Model Execution

This section is assuming that models will become more complex and therefore computationally more challenging. Models might become more complex because they cover a larger domain or because they have more detailed spatial resolutions. One area of inquiry that therefore requires advancements so that it can serve more complex models are optimization and sensitivity analysis tools. We currently explore the parameter spaces of medium complexity models in great detail – to understand the location of the best parameter sets or to understand dominant controls. However, we regularly find that the most complex models are calibrated or analysed using semi-manual or manual strategies, which suggests that there is a mismatch between the models most in need of powerful tools, and the tools we have at our disposal. Most of our currently available tools fall down when confronted with very large problems, i.e. large parameter spaces.

Computational demands can be reduced if we better understand which (modelled) processes are dominant (at particular times or in particular parts of the model domain) – therefore allowing us to search reduced parameter spaces rather than the very large parameter spaces of these models. Our sensitivity analysis methods are not yet particularly good to understand highly interacting parameter spaces though, which is what we typically encounter in complex models. There is also still a lack of how we effectively merge process understanding with optimization/sensitivity analysis to derive approaches tailored to our complex hydrologic models.

We agree with this sentiment. The original paper discussed both surrogate models and computationally frugal model analysis methods, though did not go into great detail. We have revised the paper to highlight the need for model analysis as a key path forward for the community.

> *We need to advance methods for model analysis, especially for complex models. As mentioned above, analysis of complex models is possible by both (a) developing surrogate models, i.e., models that emulate the behavior of complex models and run very quickly [Razavi et al. 2012]; (b) applying computationally frugal model analysis methods that require a fewer number of model simulations [Rakovec et al. 2014; Hill et al. 2015]; and (c) developing multi-scale methods that provide insight into finer time-space scale behavior at only the cost of coarser time-space analysis [Samaniego et al. 2010; Rakovec et al. 2015]. These advances in model analysis are important because complex models are typically calibrated or analyzed using semi-manual or manual strategies, largely because of their immense computational cost (it is only possible to run a handful of simulations). We have very little insight process/parameter dominance and process/parameter interactions in very complex models, and such information is desperately needed in order to inform meaningful parameter estimation strategies.*